PCT/GB 99/04352

Europäisches Patentamt

European Patent Office

Office européen des brevets

ZJV

GB99/4352

REC'D 23 MAR 2000

WIPO PCT

Bescheinigung Certificate Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

98310567.7

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

BEST AVAILABLE COPY

Der Präsident des Europäischen Patentamts:
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

I.L.C. HATTEN-HECKMAN

DEN HAAG, DEN
THE HAGUE, 01/03/00
LA HAYE, LE

EPA/EPO/OEB Form 1014 - 02.91

Europäisches Patentamt

European Patent Office

Office européen des brevets

Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr
Application no
Demande n° 98310567.7

Anmeldetag
Date of filing
Date de dépôt 22/12/98

Anmelder
Applicant(s)
Demandeur(s)
ISIS INNOVATIONS LIMITED
Oxford, Oxfordshire OX1 3UB
UNITED KINGDOM

Bezeichnung der Erfindung
Title of the invention
Titre de l'invention
Method of sequence identification

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat State Pays	Tag Date Date	Aktenzeichen File no Numéro de dépôt

Internationale Patentklassifikation
International Patent classification
Classification internationale des brevets
C12Q1/68

Am Anmeldetag benannte Vertragstaaten
Contracting states designated at date of filing AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE
Etats contractants désignés lors du depôt

Bemerkungen
Remarks
Remarques

METHOD OF SEQUENCE IDENTIFICATION

Introduction

There is an urgent need to identify and develop new families of antibiotics to combat the continuing threat of antibiotic resistance. Most major pharmaceutical companies are using the following general approach to identify new antibiotics.

1. Target identification. Genes that appear to be essential for bacterial viability are first identified. Most often, this is based on inviability of the cells containing a form of the gene that has been inactivated. Information from genomic DNA sequences is used to identify genes that are conserved in bacteria but not in humans. The targeting of such genes is thought to be more likely to provide inhibitors that will be specific to bacterial cells - a crucial feature for efficacy as an antibiotic.

2. Assay development A suitable means of screening for inhibition of the function of the target gene is then devised. This can be either an *in vitro*, biochemical assay, or a whole cell assay.

3. High throughput screening The assay is configured in such a way that a very large number of chemical compounds can be screened for activity against the target function. Many pharmaceutical companies have access to very large collections of compounds that can be screened. Compounds detected in this way provide lead molecules from which antibiotics can potentially be derived.

Most pharmaceutical companies now have very good resources with which to underpin activities 1 and 3. However, there is a major bottleneck in the development of assays, especially for the large number of genes of unknown function that are potentially good targets.

Most important functions in living cells, especially bacteria

and lower eukaryotes, are tightly regulated. Feedback mechanisms often ensure that products are made only in amounts sufficient to fulfil their functions. There are numerous examples of feedback regulation at the level of transcription or translation; for example, transcriptional attenuation to control tRNA synthetase (Henkin, 1994) and ribonucleotide precursor synthesis (Lu *et al*, 1996); DNA supercoiling controls the promoter for DNA gyrase (Menzel & Gellert, 1983). The same is likely to occur at the promoters of many other genes, including genes for cell division, DNA replication, etc.

The present invention makes use of this feedback regulation, hereinafter sometimes called autoregulation, to provide whole cell assays for screening compounds for antibiotic or other biological activity. It may be convenient for this purpose to make use (to effect expression of a reporter gene) of a promoter for the target gene. Alternatively other regulatory sequences, e.g. promoters of other genes of known or unknown function, may be used for the purpose. It is another object of this invention to provide a means of identifying such regulatory sequences.

The Invention

In one aspect the invention provides cells of an organism suitable for screening compounds for biological activity, which cells contain a chromosome including:

a) a target gene whose expression or activity is subject to a feedback mechanism, and

b) an artificially introduced reporter gene under the control of a regulatory sequence associated with the said feedback mechanism,

whereby a reduction of synthesis or activity of a target gene expression product is associated with an increase in the expression of the reporter gene.

In another aspect the invention provides a method of

screening compounds for biological activity, which method comprises incubating the compounds with aliquots of the cells as defined, and observing the level of expression of the reporter gene.

In another aspect, the invention provides compounds e.g. antibiotics identified by the method; and use of the compounds so identified to treat, e.g. kill or inhibit the growth of, bacteria.

The organism is preferably a bacterium e.g. *Bacillus subtilis*, although other prokaryotes and even simple eukaryotes such as yeasts are envisaged. It may be convenient to use an organism whose genome is sufficiently small that it can be chopped up into random fragments of convenient size which constitute a library of potential regulatory sequences, with the number of potential regulatory sequences of the library not being impractically large. This is a property possessed by bacteria and possibly also by yeasts.

A target gene is any gene of known or unknown function whose expression *in vivo* is subject to a feedback mechanism. The nature of the feedback mechanism is not critical; it may operate at the RNA level or at the protein level and may involve transcription and/or translation of the target gene and/or activity of a target gene expression product. The target gene will generally encode a protein that is essential for viability and well conserved in bacteria or other micro-organisms but absent from or highly dissimilar in higher organisms e.g. humans. These include proteins involved in the following process:

Cell wall (peptidoglycan) precursor synthesis (including the products of the genes: *murA,B,C,D,E,F,G,Z, dal, ddlA*);

Teichoc acid synthesis (*tagA,B,C,D,E,F,G,H, gtaB*);

DNA replication (*dnaA, gyrA,B, topA,B, lig[yerG]*);

RNA synthesis (*sigA, rpoA, rpoB, rpoC*);

Cell division (*ftsA,L,Z, divIB, divIVA*);

Chromosome segregation (*spoIIIE, spoOJ codV, ripX*);

Translation (*infA,B,C,fmt, efp*);

Miscellaneous (*obg, lgt*);

Essential genes of unknown function discovered by the inventor (*yjbN, yloQ*).

Examples of target genes are discussed in more detail below.

When expression of the target gene or synthesis or activity of a target gene expression product is altered (the means by which that alteration is effected being unimportant for present purposes), then a feedback mechanism may operate in the cell to compensate for the alteration. The feedback mechanism may operate on a regulatory sequence e.g. a promoter of the target gene. Thus for example a reduction of synthesis or activity of a target gene expression product may be associated, via the feedback mechanism, with an increase in activity of the promoter of the target gene; and advantage can be taken of this by artificially introducing a reporter gene under the control of the said promoter of the target gene.

Alternatively or additionally, a reduction of synthesis or activity of a target gene expression product may be associated with an alteration, generally an increase, in the activity of some other regulatory sequence, e.g. a promoter of some other gene. In the current state of knowledge of even the simplest organisms, it is generally not possible to predict what regulatory sequences (other than a promoter of the target gene) will be involved in a feedback mechanism. It may be advantageous to introduce a reporter gene under the control of such an extraneous regulatory sequence (i.e. a regulatory sequence that is not a promoter of or otherwise obviously associated with target gene). This invention provides methods of identifying potential extraneous regulatory sequences of this kind. Once such an extraneous regulatory sequence has been identified, it is a simple matter to introduce a reporter gene under its control; for example it is possible to genetically modify the chromosome of the cell by

introducing a genetic construct containing the regulatory sequence functionally fused to the reporter gene.

The cells of this invention thus contain an artificially introduced reporter gene under the control of a regulatory sequence associated with the target gene. Preferably the cells also contain a different artificially introduced reporter gene under the control of a regulatory sequence not associated with the said feedback mechanism.

The following are envisaged as preferred target genes for use in *B. subtilis* or other bacteria.

DnaA

The *dnaA* gene is widely conserved among bacterial species and in *E. coli* and *B. subtilis* at least, it is essential, being required for initiation of DNA replication (although the Cyanobacterial gene was recently reported to be non-essential; Richter *et al.*, 1998, *J. Bacteriol.* **180**, 4946-4949). Eukaryotes, in contrast, use a quite different mechanism to initiate chromosome replication, so agents that specifically inhibit DnaA function should have selective toxicity against bacteria. Accumulation of DnaA during the cell cycle is thought to result in gradual filling of binding sites ("DnaA boxes") in the *oriC* region. Eventually, enough DnaA accumulates to allow formation of an initiation complex that leads to the initiation of bidirectional chromosome replication. In *B. subtilis* at least, the *dnaA* gene is monocistronic and located right next to *oriC* (Moriya *et al.*, 1988, EMBO J. **7**, 2911-2917). There is a long upstream regulatory region containing several DnaA boxes. Two of these appear to overlap the promoter region of the gene. There is well-documented evidence in both *E. coli* and *B. subtilis* that *DnaA* negatively autoregulates, so depletion of DnaA is compensated for by increased transcription from its promoter. Preliminary experiments are in progress to show that depletion of DnaA protein (by artificially repressing its gene) result in increased transcription

from the *dnaA* promoter. If this works, reporter genes coupled to the *dnaA* promoter should provide a convenient means of screening for inhibitors of this excellent antibiotic target. Promoters to genes uninvolved with DNA replication (e.g. the P_{xyl} promoter, induced by the presence of xylose) should be either unaffected by inhibitors of DnaA, or else gradually shut down, as the cells begin to die through loss of their DNA. It is possible that the arrest in initiation of DNA replication caused by depletion or inhibition of DnaA will result in repression of genes involved in other steps in DNA replication. The promoters of such genes might provide better control promoters to ensure the specificity of inhibitors against DnaA, rather against DNA replication in general. A dual reporter system based on the P_{dnaA} and possibly P_{xyl} promoters should provide a sensitive and specific whole cell assay for inhibitors of DnaA function.

Dal

Bacterial cell walls are essential stress-bearing structures made of peptidoglycan (PG). A wall is present in virtually all eubacteria. It is composed of long glycan (amino sugar) polymers crosslinked by short peptides. The peptides are unusual in that they contain amino acids in the D-isomeric state. There is no structure equivalent to the cell wall in higher eukaryotes and D-amino acids are generally absent. Many important antibiotics, including penicillins, cephalosporins and vancomycin act on PG synthesis. D-alanine, which is universally found in PG, is obtained by bacteria either from their growth medium or by the action of an enzyme, D-alanine racemase, encoded by the *dal* gene, which converts L-alanine to the D-form. *dal* is a very highly conserved gene in eubacteria but absent from eukaryotes.

Tests are in progress to determine whether the *dal* promoter is autoregulated; its promoter is expected to be up-regulated in response to depletion of the Dal protein. If this is the case, the promoter can be

used to screen for inhibitors of Dal. P_{xyl} can again be used as a control, although there might be better promoters (e.g. one responding negatively to increases in the substrate used by Dal).

InfC

The process of translation initiation is highly conserved in bacteria but differs fundamentally from eukaryotes. Thus, initiation codons are defined by a "ribosome-binding site" (RBS), which lies close to the initiation codon, which is usually AUG but sometimes UUG or GUG.

In humans, the first AUG in the mRNA tends to be used, irrespective of the context - there is no equivalent of the RBS. Translation initiation factor 3 of bacteria is highly conserved in bacteria but apparently absent from eukaryotes as might be expected. It seems to be required for specificity of initiation at start codons. In both *B. subtilis* and *E. coli*, the initiator codon of the *infC* gene is highly unusual - AUU. Probably, this codon can only be used to initiate translation and thus make more InfC protein, when the factor is limiting, allowing the ribosome to initiate less stringently. Experiments with *E. coli* support this (reviewed by Grunberg-Manago, 1996, In *Escherichia coli* and *Salmonella typhimurium: cellular and molecular biology,* eds Neidhardt *et al.*, ASM Press, Washington D.C. pp 1432-1457). A highly specific whole cell assay for inhibitors of InfC is envisaged based on a use of a pair of reporter genes. Both will carry a promoter driving transcription of the *infC* translation initiation region; in one case, the reporter gene will be fused in frame with the beginning of the *infC* coding region, including the AUU initiation codon; the other will have a different reporter gene fused to an identical leader except with AUG as initiator. The first reporter should induce when InfC is depleted or inhibited. The latter will provide a control essentially unresponsive to changes in the levels of InfC.

A difficulty in validating this assay lies in the presence of

likely essential genes, *rpmI* and *rplT* (ribosomal protein genes), downstream from *infC*. It may be necessary to inactivate *infC* with integrative plasmid pMUTIN4, for example, driving downstream genes with IPTG, then put *infC* under inducible control elsewhere in the chromosome, e.g. with P_{xyl} at *amyE*. (Vagner, V., Dervyn, E and Ehrlich, S.D. (1998) Microbiology **144**, 3097-3104).

Fmt

Methionyl-tRNA formyltransferase (Fmt) carries out the final step in synthesis of the initiator tRNA in bacteria. Bacterial cells use a special initiator tRNA, $tRNA_{Fmt}$, specifically at start codons. Eukaryotic cells use the normal tRNA for methionine at both initiator and internal codons and so have no need of Fmt. Disruption of the *fmt* gene of *E.coli* (at least) causes a severe growth defect (Guillon, J.-M., Mechulam, Y., Schmitter, J.-M., Blanquet, S., Fayat, G. (1992) J. Bacteriol, **174**, 4294-4301), so its product is likely to be a good target for antibiotics. In the gene sequence, there is an unusual RBS with an ATG right next to it, followed by an Ile codon and a stop. The correct initiator AUG follows. In the presence of excess Fmt, binding is expected to occur at the inappropriate upstream AUG and this might block use of the correct start codon. This could be tested with wild type and mutant initiator regions fused in frame to *lacZ*, as for *infC*. It is expected that these reporters will respond appropriately to depletion of Fmt, and will provide a specific assay for inhibitors of Fmt.

When the regulatory sequence is not a promoter of the target gene, then the invention also provides a method of identifying a regulatory sequence, which method comprises the steps of:-

i) selecting a target gene in a chromosome of an organism wherein expression of the target gene is subject to a feedback mechanism,

ii) altering the synthesis or activity of an expression product of the target gene, and

iii) observing a corresponding change in activity of a regulatory sequence associated with the said feedback mechanism,

wherein there is introduced into the chromosome of the organism a reporter gene under the control of the regulatory sequence, and step iii) is performed by observing a corresponding change in the expression of the reporter gene.

Step ii) of that method preferably involves controlling the expression of the target gene through the use of a genetic construct containing a known repressor or promoter sequence so as to modulate a level of an expression product via exposure of the cells to an external chemical or physical inducing factor; or isolating a conditional mutation in the gene, which allows activity of its product only under certain conditions, such as low or high temperature.

In one method there is introduced into the chromosome of the organisation a reporter gene under the control of the regulatory sequence, and step iii) is performed by observing a corresponding change in the expression of the reporter gene. Preferably there is provided a library of constructs of potential regulatory sequences associated with the said feedback mechanism, each potential regulatory sequence fused to a reporter gene. Preferably iii) is performed by genetically modifying aliquots of cells of the organism by introducing into the chromosome of the cells of each aliquot a different construct of the library. The potential regulatory sequences are preferably DNA fragments of approximately 50 – 1000 bp of the genome of the organism. A library would preferably contain at least 10 different constructs. Preferably the DNA fragments of the library comprise the entire genome of the organism.

Another method involves providing an array of DNA sequences of genes of the organism. Preferably the array is of DNA sequences taken from each of the genes of the organism and provided in spots at spaced locations on a surface of a support. Cells of the organism

are incubated under conditions which permit expression of the target gene. RNA is recovered and if desired converted to cDNA. Either the RNA or the cDNA is applied to the array under hybridisation conditions, and a hybridisation pattern is noted. Another aliquot of cells of the organism is incubated under conditions to alter or prevent expression of the target gene. RNA is recovered and RNA or cDNA is applied to the array under hybridisation conditions in order to generate a hybridisation pattern different from the previous one. The differences are noted and are indicative of genes and their regulatory sequences which are involved in a feedback mechanism associated with the target gene. Such gene expression arrays have been described in the literature and used to investigate feedback mechanisms of micro-organisms, but not, it is believed, for the purpose of identifying regulatory sequences for use in screening for antibiotics. (DeRisi, J L *et al*, 1997).

In yet another method, step iii) is performed by recovering proteins from cells of the organism, separating the recovered proteins and observing a corresponding change in concentration of at least one individual protein. Thus a first aliquot of cells of the organism is cultured under conditions to permit expression of the target gene. Proteins in the cell are recovered and separated e.g. by use of a two dimensional electrophoresis gel. This technique is well known in the literature (Anderson, N L *et al*, 1998) and gives rise to a pattern where, at least in the case of *Bacillus* species, the identity of each protein is known. Then another aliquot of the cells is incubated under conditions to prevent expression of the target gene. Recovered proteins are separated by electrophoresis to produce a pattern different from the first. The differences are observed and are indicative of genes and regulatory sequences that are involved in a feedback mechanism associated with the target gene.

The first of these three methods involves a substantial capital

investment to create a library of constructs of potential regulatory sequences each fused to a reporter gene; but it should be effective to identify any or all regulatory sequences involved in the feedback mechanism associated with the target gene. The second method using gene expression arrays, avoids the expense of creation of a library, and should be effective to pick up regulatory sequences involved in feedback mechanisms where transcription is altered. The third method, using "proteomics" is relatively simple and cheap, but may not be effective to pick up regulatory sequences where the feedback mechanism involves mRNA synthesis or protein stability or activity.

Use of these methods permits regulatory sequences, involved in feedback regulation of a target gene but which are not promoters of the target gene, to be identified. Once this has been done, it is a simple matter to construct cells as defined, containing an exogenous reporter gene under the control of the regulatory sequence, and to use these cells in a whole cell assay for screening compounds for antibiotic or other biological activity.

Detailed Description

For any given target gene, there is thus a good chance that somewhere in the genome there will be a regulatory sequence the activity of which is enhanced or reduced by lack of the target function. In *B. subtilis* it is very straightforward to turn off any target gene by making genetic constructions in which the gene can be repressed (e.g. by use of the IPTG-inducible P_{spac} promoter (Yansura & Henner, 1984), or the xylose-inducible promoter, P_{xyl} (Feucht *et al*, 1996)). When the inducer compound is taken away, expression of the gene is blocked and the target function is depleted from the cell. As discussed above, this is likely to result in specific induction of one or more genes. To find the promoter or regulatory sequences for such a gene, a large (> 10,000) random collection of short

(approximately 50 to 1,000 bp) DNA sequences from the chromosome of *B. subtilis* or a related organism are fused to an appropriate reporter gene (e.g. *lacZ).* Methods known in the art are available to achieve this (reviewed by Errington, 1990). Each member of the collection of fusions is then introduced into a cell of a strain of *B. subtilis* that had been genetically engineered to allow depletion of the target function. A regulatory sequence exhibiting the desirable property of modulating reporter gene expression upon depletion of the target function would provide a means of screening for compounds which are specific modulators of the target gene function. For example, if the gene encoding one of the subunits of DNA gyrase (*gyrA*) is depleted, the resultant change in DNA supercoiling in the cell should result in increased expression from the promoter of the *gyrA* gene. Some random fragments of DNA (e.g. one containing the *gyrA* promoter itself), when fused to a reporter gene and introduced into cells in which gyrase has been depleted, would give rise to increased reporter activity. Isolating the gyrA promoter fragment (or some similarly behaving regulatory sequence), and fusing it to a reporter gene, gives a genetic construct that would respond, with increased expression, to inhibition of DNA gyrase by chemical inhibitors (i.e. potential antibiotics) (Menzel & Gellert, 1983). In any operational screen, it is preferable to include a second reporter gene, encoding a non-gyrase-dependent promoter fragment, to control for non-specific inhibitors of gene expression. In principle, it might sometimes be possible to find, in the random collection, a DNA sequence which, when fused to the reporter and depleted for e.g. the *gyrA* product, would respond with decreased expression. Such a reporter would provide a control reporter that would most likely increase the specificity of the assay for inhibitors of DNA gyrase.

A high throughput screen runs in the following way. A strain of *B. subtilis* is constructed containing two reporter genes encoding enzymes or proteins that can be detected, such as β-galactosidase and β-

glucuronidase. Reporter genes capable of being expressed in *Bacillus* species are well known and documented in the literature. Reporter genes are preferably chosen so that their products can be readily assayed simultaneously. Green fluorescent protein has the advantage that its intrinsic fluorescence allows the protein to be assayed by direct fluorimetric measurement. *LacZ* has been used for more than 10 years with great success in *B. subtilis* and there is a range of useful substrates that generate coloured or fluorescent products upon hydrolysis by β-galactosidase. The *uid* gene of *E. coli* has recently been harnessed for similar purposes, and the range of substrates available for the gene product, β-glucuronidase is similar to that for β-galactosidase. Two different fluorogenic substrates may be used to assay the activities of the two reporters simultaneously in a single reaction. One reporter is fused to a sequence that causes increased expression in the absence of the desired function (e.g. gyrase); the other to a sequence that gives unchanged or decreased expression. The dual reporter strain is grown in an appropriate medium, dispensed into a vessel allowing large scale screening, such as the wells of microtitre plates. Each well would contain at least one test compound. The cells would be incubated for an appropriate time period (preferably two or more cell doublings), after which (if necessary, depending on the reporters used) a reaction cocktail allowing simultaneous assay of the two reporter enzymes would be added. The assays might be based on enzyme substrates giving e.g. chemiluminescent or fluorescent or coloured products. For fluorogenic substrates the presence of either or both enzymes may be detected simultaneously by a fluorimeter set to receive two different appropriate wavelengths. In the example given, a positive response, in terms of inhibition of DNA gyrase, would be indicated by increased activity of β-galactosidase and decreased β-glucuronidase. Compounds eliciting such a response would be potential antimicrobial agents with DNA gyrase

as the likely target.

Alternatively, the chemicals could be individually spotted onto a lawn of cells plated on a growth medium containing colorigenic or fluorogenic substrates. Chemicals eliciting an appropriate response would be detectable by their enhancement or inhibition of enzyme activity in the vicinity of the spot.

In principle, there are many essential genes in bacteria that are likely to be sufficiently different from mammalian cells to provide the selective toxicity needed for a useful antibiotic. Moreover, it is well known that mutations affecting central metabolic functions, such as nucleotide precursor synthesis, lead to attenuation of virulence, because they perturb the ability of the pathogen to grow in host cells or tissues. So this approach may provide not only antibiotics but also drugs that attenuate virulence and which will act synergistically with antibiotics. The strategy is especially important because it could be applied to any gene of interest, even though the precise function of the gene is not known.

The general strategy may be applicable to other organisms. Yeast genetics should be facile enough to allow it to be used. This opens up the possibility of identifying compounds that inhibit specific eukaryotic functions; at least those that are conserved in yeast. The yeast strain may first have its endogenous gene replaced with the equivalent human gene to make the screen more direct and specific.

Example 1

A plasmid (pAT1) containing the N-terminal-coding region of the *gyrA* gene of *B. subtilis* is transformed into *B. subtilis* so as to create strain XXX. Integration of this plasmid into the chromosome by homologous recombination results in partial duplication of the *gyrA* gene. The functional copy of *gyrA* that remains has been placed under the control of the repressible P_{spac} promoter, so that removal of IPTG from a

growing culture results in depletion of GyrA protein. GyrA is one subunit of the essential DNA gyrase protein, so the culture soon stops growing. The plasmid also places a *lacZ* reporter gene under the control of the natural promoter for *gyrA*. The level of expression of this reporter is greatly increased by depletion of GyrA, as compared with the control culture in which IPTG remains present. In principle, the promoter for the *gyrA* gene therefore provides a means of detecting inhibitors of DNA gyrase, since these should result in increased expression of the reporter gene, compared with control reporters. This strain also contains a gene encoding a modified copy of the green fluorescent protein (GFP; Chalfie *et al.*, 1994, *Science* **263**, 802-805). This gene, which has been placed at the *amyE* locus in the chromosome, is expressed constitutively from the P_{xyl} promoter (provided that xylose is present in the medium). Measurement of the fluorescence due to the GFP reporter reveals that depletion of GyrA reduces expression of this reporter, rather than enhancing expression, as seen for the *gyrA* promoter. P_{xyl}-*gfp* thus provides an example of a control reporter gene, which would help to ensure that any chemical compounds identified act on DNA gyrase (or hypothetically an as yet unidentified functionally related protein, inhibition of which confers similar effects on cell physiology). To test this, known inhibitors of DNA gyrase (nalidixic acid and oxolinic acid) are used to treat cells of strain XXX grown in the presence of IPTG. At appropriate concentrations, both inhibitors cause increased expression of the *lacZ* reporter gene, whereas the GFP reporter is reduced.

REFERENCES

Menzel, R. & Gellert, M. (1983) Regulation of the genes for *E. coli* DNA gyrase: homeostatic control of DNA supercoiling. Cell 34, 105-113.

Yansura, D.G. & Henner, D.J. (1984) Use of the *Escherichia coli* lac repressor and operator to control gene expression in *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA 81, 439-443.

Errington,J. (1990) Gene cloning techniques. In: Molecular biological methods for Bacillus. (eds C.R. Harwood & S.M. Cutting), pp. 175-220. Chichester, U.K.: Wiley.

Henkin, T.M. (1994) tRNA-directed transcription antitermination. Mol. Microbiol. 13, 381-387.

Feucht, A., Magnin, T., Yudkin, M.D. & Errington, J. (1996) Bifunctional protein required for asymmetric cell division and cell-specific transcription in *Bacillus subtilis*. Genes Devel. 10, 794-803.

Lu, Y., Turner, R.J. & Switzer, R.L. (1996) Function of RNA secondary structures in transcriptional attenuation of the *Bacillus subtilis* pyr operon. Proc. Natl. Acad. Sci. USA 93, 14462-14467.

DeRisi *et al* , 1997, Exploring the metabolic and genetic control of gene expression on a genomic scale, Science, 278, 680-686.

Anderson, N L, *et al*, Proteome and Proteomics: New Technologies, new concepts and new words. Electrophoresis, 1998, August; 19 (11): 1853-61.

CLAIMS

1. Cells of an organism suitable for screening compounds for biological activity, which cells contain a chromosome including:

a) a target gene whose expression or activity is subject to a feedback mechanism, and

b) an artificially introduced reporter gene under the control of a regulatory sequence associated with the said feedback mechanism,

whereby a reduction of synthesis or activity of a target gene expression product is associated with an increase in the expression of the reporter gene.

2. The cells of claim 1, wherein the target gene is selected from *gyrA*, *dnaA* or other DNA replication gene, *dal*, *infA*, *infB*, *infC* or *fmt*.

3. The cells of claim 1 or claim 2, wherein the cells also contain a different artificially introduced reporter gene under the control of a regulatory sequence not associated with the said feedback mechanism.

4. A method of screening compounds in order to identify one or more of the compounds as having biological activity, which method comprises incubating each compound with an aliquot of the cells of any one of claims 1 to 3, and observing the level of expression of the at least one reporter gene.

5. A compound identified by the method of claim 4.

6. The compound of claim 5 which is an antibiotic.

7. A method of killing or inhibiting the growth of bacteria, which method comprises applying to the bacteria the compound of claim 6.

8. A method of identifying a regulatory sequence, which method comprises the steps of:-

i) selecting a target gene in a chromosome of an organism wherein expression of the target gene is subject to a feedback mechanism,

ii) altering the synthesis or activity of an expression product of the target gene, and

iii) observing a corresponding change in activity of a regulatory sequence associated with the said feedback mechanism,

wherein there is introduced into the chromosome of the organism a reporter gene under the control of the regulatory sequence. and step iii) is performed by observing a corresponding change in the expression of the reporter gene.

9. A method of identifying a regulatory sequence, which method comprises the steps of:-

i) selecting a target gene in a chromosome of an organism wherein expression of the target gene is subject to a feedback mechanism,

ii) altering the synthesis or activity of an expression product of the target gene, and

iii) observing a corresponding change in activity of a regulatory sequence associated with the said feedback mechanism,

wherein there is provided an array of DNA sequences of genes of the organism, and step iii) is performed by recovering RNA from cells of the organism, applying the RNA or cDNA to the array under hybridisation conditions, and observing a corresponding change in a pattern of hybridisation.

10. A method of identifying a regulatory sequence, which method comprises the steps of:-

i) selecting a target gene in a chromosome of an organism wherein expression of the target gene is subject to a feedback mechanism,

ii) altering the synthesis or activity of an expression product of the target gene, and

iii) observing a corresponding change in activity of a regulatory sequence associated with the said feedback mechanism,

wherein step iii) is performed by recovering proteins from cells of the organism, separating the recovered proteins by "proteomics" and observing a corresponding change in concentration of at least one individual protein.

11. The method of any one of claims 8, 9 and 10, comprising an additional step of:

iv) modifying cells of the organism by placing an artificially introduced reporter gene under control of the regulatory sequence observed in step iii).

ABSTRACT

METHOD OF SEQUENCE IDENTIFICATION

Cells of *Bacillus subtilis* or other micro-organism suitable for screening compounds for antibiotic or other biological activity contain a chromosome including: a target gene whose expression is subject to a feedback mechanism; and an artificially introduced reporter gene under the control of a regulatory sequence associated with the feedback mechanism, whereby a reduction of synthesis or activity of an expression product of the target gene is associated with an increase in the expression of the reporter gene. A method of screening compounds for antibiotic or other biological activity comprises incubating the compounds with aliquots of the cells as defined and observing the level of expression of the reporter gene. Methods of identifying suitable regulatory sequences which are not promoters of the target gene, are also described.